Lehrstuhl Angewandte Informatik IV
Datenbanken und Informationssysteme
Prof. Dr.-Ing. Stefan Jablonski

Institut für Angewandte Informatik
Fakultät für Mathematik, Physik und Informatik
Universität Bayreuth

Seminar

Vorname Nachname

August 26, 2015

Version: Draft / Final

Universität Bayreuth

Fakultät Mathematik, Physik, Informatik
Institut für Informatik
Lehrstuhl für Angewandte Informatik IV

Titel / Topic

Seminar

Vorname Nachname

1. Reviewer **Prof. Dr.-Ing. Stefan Jablonski**
Fakultät Mathematik, Physik, Informatik
Universität Bayreuth

2. Reviewer **Dr. Stefan Schönig**
Fakultät Mathematik, Physik, Informatik
Universität Bayreuth

Supervisors Stefan Schönig and Lars Ackermann

August 26, 2015

Vorname Nachname
Seminar
Titel / Topic, August 26, 2015
Reviewers: Prof. Dr.-Ing. Stefan Jablonski and Dr. Stefan Schönig
Supervisors: Stefan Schönig and Lars Ackermann

Universität Bayreuth
Lehrstuhl für Angewandte Informatik IV
Institut für Informatik
Fakultät Mathematik, Physik, Informatik
Universitätsstrasse 30
95447 Bayreuth
Germany

Abstract

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Abstract (different language)

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

Acknowledgement

Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

This is the second paragraph. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language. Hello, here is some text without a meaning. This text should show what a printed text will look like at this place. If you read this text, you will get no information. Really? Is there no information? Is there a difference between this text and some nonsense like “Huardest gefburn”? Kjift – not at all! A blind text like this gives you information about the selected font, how the letters are written and an impression of the look. This text should contain all letters of the alphabet and it should be written in of the original language. There is no need for special content, but the length of words should match the language.

i really wanna stay at your house

Contents

1 Introduction

1.1 Motivation

Machine learning (ML) has become a pivotal tool in decision-making processes across numerous domains, including healthcare, finance and hiring. However, as these models increasingly influence critical decisions, questions about their fairness and the ethical implications of their use have come to the forefront. Fairness in ML concerns the equitable treatment of individuals or groups, particularly in the presence of sensitive attributes such as gender, age, race, or socioeconomic status. These attributes often correlate with historical inequalities or systemic biases embedded in the data. ML models, designed to optimize accuracy, learn patterns from this data and may inadvertently exploit these unfair patterns to make predictions. While leveraging such patterns can improve predictive performance, it also risks perpetuating or even amplifying existing inequities. A notable example of this occurred when Amazon's AI recruiting tool, which was designed to assist in hiring decisions, exhibited significant gender bias, favoring male candidates due to biases in historical hiring data, leading to its eventual scrapping. [Cha23]

These challenges of fairness and bias are not confined to traditional applications but also extend to the field of predictive business process monitoring (PBPM). Here, ML models play an integral role in making decisions, whether by predicting outcomes, allocating resources, or streamlining operations. Addressing fairness in this context requires careful consideration of bias's dual nature. While biases can lead to discrimination, not all bias is inherently harmful. In some cases, certain biases may be necessary for the model to achieve its intended purpose. For instance, sensitive attributes like gender can carry significant information relevant to specific contexts. In the domain of healthcare, gender differences in biological and hormonal factors are crucial for determining effective treatments and drug prescriptions. However, the same attribute might lead to discriminatory outcomes in unrelated contexts, such as predicting whether a patient's request for treatment will be approved. This duality highlights the complexity of fairness in ML: biases that are helpful in one scenario can become harmful in another.

A significant obstacle in addressing this issue is the inherent opacity of ML models. Many models, especially complex ones such as deep neural networks, operate as "black boxes" that produce predictions without providing insight into how those predictions are made. This lack of interpretability makes it challenging for human stakeholders and domain experts to identify whether a model's reliance on a sensitive attribute aligns with ethical and operational goals. This presents a dilemma: either leaving potentially harmful biases unchecked to avoid significantly compromising the model's predictive performance, or removing sensitive attributes entirely from the model's input. The latter approach, while intended to prevent discrimination, can lead to suboptimal predictions, especially when the sensitive attributes carry critical information relevant to the task.

1.2 Problem Statement

To address this challenge, this thesis proposes an approach based on knowledge distillation. Knowledge distillation involves transferring the encapsulated knowledge from a complex model to a simpler, more interpretable representation. By applying this technique, the inner workings of a predictive ML model can be visualized and understood by human stakeholders and domain experts, which in turn enables the identification of inherent biases in the model. Specifically for the proposed approach, the distilled representation is modified to remove the unwanted bias by adjusting the way sensitive attributes influence the representation's decision-making process. This modified version of the representation is then used to relabel the training data, creating a refined dataset that reflects the desired fairness characteristics. Finally, the original model is fine-tuned using this new, bias-adjusted training data, allowing it to learn from the corrected representation and ultimately make more equitable predictions.

The goal of this approach is twofold: to make the ML model fairer by eliminating the biases that result in unfair treatment and to maintain its predictive accuracy by preserving helpful ones. By balancing these objectives, the proposed methodology seeks to create models that are not only effective but also aligned with given ethical standards and societal expectations.

1.3 Thesis Outline

Building on the problem statement, the second chapter delves into related work in the area of bias reduction and fairness in PBPM. This chapter provides a comprehensive review of existing research, and highlights the gaps this study aims to address.

The third chapter provides background information essential for understanding the research. It introduces predictive business process model monitoring, explains the ML models used in this thesis and discusses fairness in ML.

The fourth chapter details the methodology employed in the thesis. It describes the process of data gathering and preprocessing, with a focus on handling bias causing attributes. The chapter then explains how the initial ML model is trained, followed by the application of knowledge distillation to create an interpretable representation of the model, which can then be modified to address unwanted biases. The methodology further includes how the predictions of the modified representation can be used to finetune the model to improve fairness, while maintaining accuracy.

TODO The fifth chapter presents the evaluation of the proposed approach. This chapter defines the metrics used for evaluation and includes empirical results from multiple case studies.

TODO The final chapter concludes the thesis by summarizing the findings and discussing their implications for fairness in ML. It acknowledges the limitations of the research and offers suggestions for future work.

Related Work 2

2.1 Fairness in Predictive Business Process Monitoring

2.2 Conclusion

Current approaches to fairness in PBPM have made significant strides but remain limited by their automated and uniform treatment of bias. A shift toward stakeholder-driven, knowledge distillation-based methods presents a promising direction for achieving more balanced and context-sensitive outcomes.

3 Background

3.1 Predictive Business Process Monitoring

Process Modeling

Event Log Data

Next Activity Prediction

3.2 Machine Learning Models

Neural Networks

Decision Tree

3.3 Fairness

4 Methodology

4.1 Data Gathering

Data Simulation

Data Augmentation

4.2 Data Processing

4.3 Training the Model

4.4 Knowledge Distillation

4.5 Modification of the Decision Tree

Cutting Branches

Retraining Subtrees

4.6 Finetuning of the Model

5 Evaluation

5.1 Evaluation Metrics

Accuracy

Fairness

5.2 Evaluation Results

5.2.1 Proof of Concept

5.2.2 Cancer Screening

5.2.3 Hospital Billing

5.2.4 BPI Challenge 2012

6 Conclusion

6.1 Implications of the results

6.2 Limitations and Challenges

6.3 Future Work

Bibliography

[Cha23] Xinyu Chang. „Gender Bias in Hiring: An Analysis of the Impact of Amazon's Recruiting Algorithm". In: *Advances in Economics, Management and Political Sciences* 23 (2023), pp. 134–140 (cit. on p. 1).

List of Figures

List of Tables

Declaration

You can put your declaration here, to declare that you have completed your work solely and only with the help of the references you mentioned.

Bayreuth, August 26, 2015

Vorname Nachname